DOCUMENT RESUME

ED 124 098 HE 007 598

AUTHOR Chidambaram, T. S.
TITLE Enrollment Forecasting in an Open Admissions Environment.
INSTITUTION Federal City Coll., Washington, D.C. Office of Institutional Research.
PUB DATE 74
NOTE 54p.; Not available in hard copy due to marginal legibility of original document.

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS College Freshmen; Educational Planning; *Enrollment Projections; *Enrollment Rate; *Higher Education; Institutional Research; *Mathematical Models; *Open Enrollment; Post Secondary Education; Prediction; Predictor Variables; Public Schools; Student Enrollment
IDENTIFIERS Federal City College

ABSTRACT

Developing a model for predicting demand for freshmen requirement courses (from freshmen enrollees and from returned enrollees who failed to complete the course in their previous quarters) is the objective of the Freshmen Requirement Study, now partially completed by Federal City College. Work done so far has essentially validated an initial approach to the problem where past enrollment behavior has been taken as a predictive factor. Analysis of the data specially compiled for the study shows that the more numerous and more recent a student's enrollment has been in the past, the higher his probability of return is. The difference between the summer quarter and other quarters has been documented. A model that fits well the observed data on return probabilities has been constructed and, using it, the effects of the various components of past behavior affecting return probabilities have been measured. Future work will validate the model with more recent data, fortify it, if necessary, with other explanatory variables, and put it into operation, setting up systems for data collection analysis. Appendices discuss mathematics of probabilistic predictive models: Operationalizing Predictive Models and Fitting an Additive Model to Logarithm of Reenrollment Probabilities. (Author/JT)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED124098

Enrollment Forecasting In An Open Admissions Environment*

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

T. S. Chidambaram, Ph.D.
Ken Robert Gramza , Director

Office of Institutional Research
Federal City College

---

*To be presented at the 74th Annual Meeting of the American Educational Research Association, April 15-19, 1974, at Chicago, Illinois.

HE007998

ENROLLMENT FORECASTING IN AN OPEN ADMISSIONS ENVIRONMENT

- TO PRESENT A GENERAL SCHEME APPLICABLE FOR FORECASTING THE NUMBER OF STUDENTS RETURNING TO SCHOOL FROM PREVIOUS QUARTERS/SEMESTERS
- IN AN OPEN ADMISSIONS ENVIRONMENT STUDENTS DROP IN AND OUT AT WILL, MAKING THIS PREDICTION IMPORTANT FOR SUCH TASKS AS
  - COURSE SCHEDULING
  - ESTIMATING GRADUATE OUTPUT
  - RECRUITMENT TARGET DETERMINATION

EXHIBIT 2

- TABLE 1 SHOWS THAT ABOUT 22% OF STUDENTS ENROLLED IN ONE QUARTER DO NOT RETURN NEXT QUARTER BUT ABOUT 40% OF THESE DROPOUTS DO RETURN TO SCHOOL IN LATER QUARTERS.

  PERCENTAGE RETURNING AFTER FOUR QUARTERS OF ABSENCE IS SMALL.

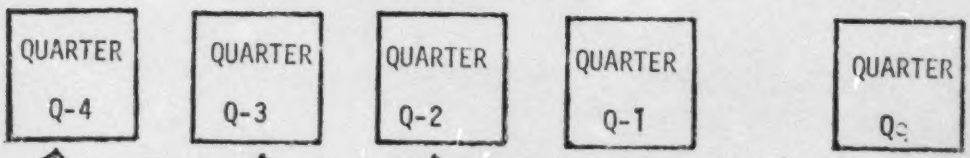
- PREVIOUS STUDIES INDICATE FUTILITY OF USING SOCIOECONOMIC VARIABLES TO PREDICT DROPOUT BEHAVIOR.

  APPROACH TAKEN HERE IS TO USE PAST ENROLLMENT BEHAVIOR ITSELF AS PREDICTOR VARIABLE. SPECIFICALLY ENROLLMENT IN PREVIOUS FOUR QUARTERS IS USED.

EXHIBIT 3

- DEFN: AFFILIATED STUDENT IN QUARTER Q IS A STUDENT WHO HAS ENROLLED IN AT LEAST ONE OF THE FOUR QUARTERS PRECEDING Q.

- EACH AFFILIATED STUDENT CAN BE ASSIGNED TO ONE OF FIFTEEN ENROLLMENT PATTERNS (REPRESENTED BY A FOUR DIGIT BINARY NUMBER) DEPENDING ON HIS ENROLLMENT PATTERN IN LAST FOUR QUARTERS.



EXAMPLE 1:	0	0	1	0	FOR A STUDENT ENROLLED ONLY IN QUARTER Q-2
2:	1	1	0	0	FOR A STUDENT ENROLLED IN QTRS Q-4 & Q-3 ONLY

EXHIBIT 4

- TABLE 2 SHOWS THE RETURN PROBABILITIES FOR STUDENTS WITH VARIOUS ENROLLMENT PATTERNS. SUMMER QUARTER IS OBVIOUSLY DIFFERENT AND SOME PATTERNS HAVE HIGHER RETURN PROBABILITIES.

- TWO-WAY ANALYSIS OF VARIANCE (TABLE 3) SHOWS SIGNIFICANT INTERACTION BETWEEN QUARTERS AND PATTERNS EVEN AFTER REMOVING SUMMER QUARTER. THIS IMPLIES THAT THE DIFFERENCE BETWEEN PATTERNS IS NOT UNIFORM FOR ALL REGULAR QUARTERS (I.E., FALL, WINTER AND SPRING). EXAMPLE: 0101 AND 0110. HENCE PATTERN DIFFERENCES ARE OBSCURED BY THIS INTERACTION.

- CONSISTANT PATTERN DIFFERENCES EMERGE (TABLE 4) ON REARRANGEMENT OF TABLE 2 DATA BY CONSIDERING PATTERNS BASED ON SUMMER QUARTERS SEPARATELY.

EXHIBIT 5

- TABLE 4 SHOWS THAT
  - ENROLLMENT IN SUMMER ALWAYS INCREASES RETURN PROBABILITY
  - THE RETURN PROBABILITY INCREASES WITH THE NUMBER OF QUARTERS ONE ATTENDS
  - MORE RECENT THE ENROLLMENT EXPERIENCE IN A REGULAR QUARTER, THE HIGHER THE RETURN PROBABILITY
- ENROLLMENT PATTERNS HAVE BEEN ARRANGED IN TABLE 4 ACCORDING TO THE ABOVE HYPOTHESES AND PRODUCES A STRICKINGLY CONSISTENT PICTURE.

EXHIBIT 6

- TO QUANTITATIVELY ESTIMATE THE EFFECT OF PREVIOUS ENROLLMENT A 17 PARAMETER MODEL OF THE FOLLOWING TYPE WAS FIT TO THE TABLE 2 DATA

$$LN(1-P) = M+B_1+B_2+B_3+B_4+E$$

WHERE

P	=	RETURN PROBABILITY IN QUARTER Q
M	=	GENERAL MEAN
B_1	=	EFFECT OF ENROLLMENT IN QUARTER Q-1
B_2	=	EFFECT OF ENROLLMENT IN QUARTER Q-2
B_3	=	EFFECT OF ENROLLMENT IN QUARTER Q-3
B_4	=	EFFECT OF ENROLLMENT IN QUARTER Q-4

- ACTUAL MODEL DISTINGUISHED BETWEEN REGULAR AND SUMMER QUARTERS AND POSTULATED SEPARATE ENROLLMENT AND DROPOUT EFFECTS IN EACH QUARTER.

EXHIBIT 7

- MODEL FIT THE OBSERVED DATA WELL EXPLAINING 90% OF VARIATION (SEE TABLE 5)

- THE LEAST SQUARE ESTIMATES OF PARAMETER VALUES WERE

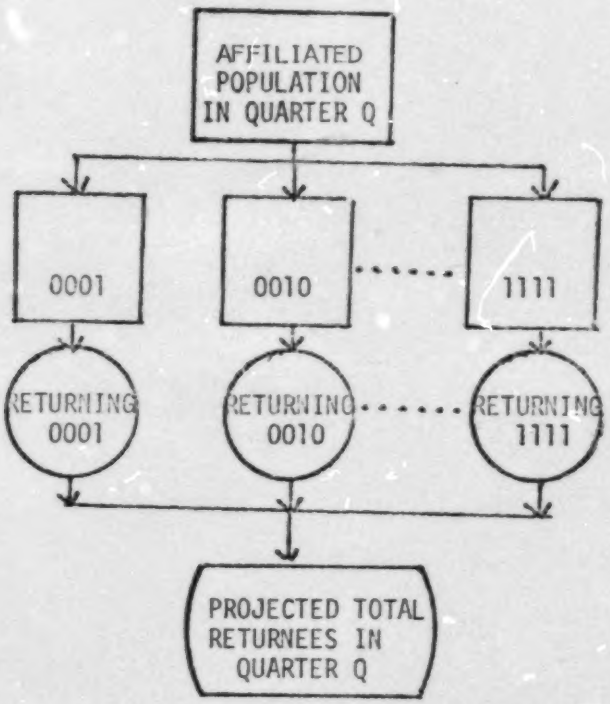
REGULAR		SUMMER	
B_1	= -1.05	B_1	= -0.84
B_2	= -0.48	B_2	= -0.49
B_3	= -0.28	B_3	= -0.16
B_4	= -0.28	B_4	= -0.15

- THE PARAMETER VALUES SUPPORT THE PREVIOUS QUALITATIVE INFERENCES.

EXHIBIT 7A

- THIS ANALYSIS SUBSTANTIATES THE VALIDITY OF USING PAST ENROLLMENT HISTORY FOR FORECASTING RETURNING STUDENTS

GENERAL FORECASTING SCHEME



BREAKDOWN INTO 15 SUBSETS BASED ON PREVIOUS ENROLLMENT

APPLY RETURN PROB APPROPRIATE FOR QUARTER Q TO EACH SUBSET

EXHIBIT 8

FORECASTING TOTAL RETURNING ENROLLEES AT FCC
DURING FALL, WINTER AND SPRING 1972-73

QUARTER

	FALL 1972-73	WINTER 1972-73	SPRING 1972-73
ACTUAL TOTAL RETURNEES	5708	5998	5158
MODEL PREDICTION STD ERROR	5968 (5895)* 283 (305)	6478 (6319) 270 (295)	6298 (6211) 267 (290)
DEVIATION DEVIATION/SE	+ 260 (+187) 0.92 (0.61)	+480 (+321) 1.8 (1.1)	+140 (+53) 0.52 (0.18)

*FIGURES IN PARANTHESES WERE OBTAINED BY USING RETURN PROBABILITIES ESTIMATED BY MODEL AS GIVEN IN TABLE 5.

EXHIBIT 9

CONCLUSIONS

- THE FORECASTING SCHEME CAN BE USED TO FORECAST RETURNING ENROLLEES FROM ANY SUBPOPULATION, E.G., ENGINEERING MAJORS.

- THE MODEL CAN BE OPERATIONALIZED VERY CONVENIENTLY AND CAN BE COMPUTERIZED.

- REQUIRES NO EXPENSIVE DATA BUT USES ONLY ROUTINE DATA COLLECTED BY INSTITUTIONS.

AN INTERIM TECHNICAL REPORT ON THE FRESHMEN REQUIREMENTS STUDY

This interim report is on the work done so far in an effort to build a predictive model for freshmen course requirements. It has essentially validated an initial approach to the problem where past enrollment behavior has been taken as a predictive factor. The analysis of the data specially complied for the study show that the more numerous and more recent a student's enrollment has been in the past, the higher his probability of return is. The difference between the summer quarter and other quarters have also been brought out. A model that fits well the observed data on return probabilities has been constructed. Using this model the effects of the various components of past behavior affecting return probabilities have been measured.

The report also discusses the general mathematics of probabilistic predictive models emphasizing the practical aspect of designing a system to operationalize a model. Finally, the work which remains to be undertaken in this study is described.

1. Introduction

The Freshmen Requirement Study (FRS) has the objective of developing a model for predicting demand for freshment requirement courses in FCC. A preliminary model was developed by Gramza, Diaz and Shore in July '72 which served as a basis for undertaking an exhaustive study starting Nov. '72. A review of the status of the study is presented below giving emphasis to both what has been done and what remains to be done.

Since the demand for freshmen courses is from fresh enrollees as well as from returning enrollees who failed to complete the course in their previous quarters, the study will have to address itself to the task of predicting a number of variables such as

i) number of new enrollees, possibly with a breakdown by credits transferred for freshmen requirement courses

ii) number of students returning from previous quarters (concentrating specifically on students who have not completed freshmen requirement courses) and

iii) number of enrolled students who are still to complete the freshmen requirement course and demand that course.

The third variable mentioned above is influenced heavily by such factors as counselling and the number of students who can be accommodated in the freshmen courses that particular quarter. However, if we are concerned with estimating demand for freshmen requirement courses with the purpose of planning enough sections, then we might justifiably ignore the third factor.

earlier, with this assumption we are effectively doing away with the necessity to consider the third factor listed above and producing a prediction that is more relevant to decision making regarding space, faculty and other resources planning. Factor (i) (namely, the new enrollees) would be an external input parameter to the predictive model.

The strategy for developing the model is as follows: Consider the problem of predicting demand for course X in Quarter i. The set of all affiliate students at start of Quarter i is the source population for all returning enrollees. Some affiliate students have passed course X and the rest have not - call the latter subset S . This is the population of interest in predicting demand for course X. The basic approach lies in partitioning S into subsets S_1, S_2, ... S_m which are mutually exclusive and exhaustive of S in such a way as to accomplish these following objectives:

(C-1) Each set s_i is a homogeneous group of affiliate students; homogeneity being used in the sense that the probability for returning to school is same for all members of the set. Perfect homogeneity can seldom be achieved in practice, since so many socio economic characteristics affect the return probability and almost any two students will have differing probabilities. However, by basing definition of the subsets on the most important of these variables, we hope to approach close to their ideal.

The demand forecast ignoring the third factor would give the number of students who need to take a freshmen requirement course in a particular quarter and planning for the courses should be based on this number. To the extent the role of counseling is to advise the students to take the courses at the earliest opportunity (subject only to the number of sections planned that quarter) the effect of counselling need not be taken into consideration separately.

Thus in this study the objective has been set as one of predicting for a given quarter and given freshmen requirement course the number of students who would need to take that course. Of the two factors (i) and (ii) listed earlier, our attention will be initially on the second one, namely the demand generated by returning enrollees. The data base and analytical techniques for predicting new enrollees are more difficult to develop while the prediction of returning enrollees can be performed with only data currently available in FCC. Further, to a certain extent FCC can regulate the number of new enrollees so that it makes sense to treat this variable as an input parameter in a predictive model rather than a variable to be itself predicted.

With these considerations as the basis, the initial scope of the Freshmen Requirements Study has been specified as one of performing the necessary statistical analysis on the available FCC data to develop a model capable of predicting the demand for freshmen requirement courses from returning enrollees, it being assumed that all those who need to take a course would indeed create a demand for the course. As explained

(C-2) FCC data should make it possible to classify a student into one of the subsets $S_1, \ldots S_m$. In other words we should not use in the predictive model any variables on which we cannot have data. Since the freshmen requirement study would construct the predictive models only on the basis of analysis of available data, this requirement should be automatically satisfied.

There are two other properties(C-3) and (C-4) which we would like the set $S_1, \ldots S_m$ to posses but these would be presented later at a more appropriate place.

Assuming we have discovered a satisfactory partitioning S_1, $S_2, \ldots S_m$ of S, the next task would be to obtain the best estimates of the parameters $r_1, r_2, \ldots r_m$ where:

r_j = the probability that a student belonging to set S_j returns to school in quarter i

It is expected that with the retention data available, this estimation can be done fairly accurately.

The prediction of number of returning students in quarter i would be then given by the expected value

$$r_1 s_1 + r_2 s_2 \ldots + r_m \qquad -(1)$$

where s_j = number of students in set S_j; j =1, .. M

The variance associated with the estimate is given by

$$r_1(1-r_1)\ s_1 + \ldots + r_m(1-r_m)s_m \ * \ -(2)$$

In order that our prediction formula (1) remains invariant over time (except possibly for predictable differences between the four quarters of a year), it is necessary that the probabilities r_j show stability over years. Again this is not often met in practice, since trends and even abrupt changes in return rates are experienced for a variety of reasons. In practice, what this means is that we have to design a system which continuously watches for trends and changes and make appropriate updating of the probabilities r_j. These and many other practical considerations in implementing a predictive model are discussed later.

The formula (2) for variance also gives us a clue as to what should be considered a satisfactory partitioning. Our aim should be to keep the variance as small as possible. From (2) it is seen that, the variance achieves the minimum value of zero, that is, estimated (1) is perfect and has no error associated with it, if each r_j is either 0 or 1.

---

* The variance formula is strictly valid only if assumption (C-1) holds. For nonhomogeneous subsets the formula would provide an upper limit, i.e., the real variance would be less.

The maximum of (2) is attained when each $r_j = 1/2$. This implies that the partitioning should be in such a way that the return probability r_j for each set is as close to zero (or one) as possible. In example we should prefer a partitioning (S_1,S_2) with probabilities (.2,.9) to a partitioning with probabilities (.3,.6). With these remarks, we formally introduce two other requirement (C-3) and (C-4) on the partitioning subset $S_1,\ldots S_m$:

(C-3) Partitioning of S into $S_1,\ldots S_m$ should be so done that the associated return probabilities $r_1,\ldots r_m$ are stable over time, [There is reason to believe that if (C-1) is satisfied then (C-3) is also likely to be satisfied].

(C-4) Partitioning of S into $S_1,\ldots S_m$ should be so done that the associated return probabilites $r_1,\ldots r_m$ are all close to zero or one.

This background discussion on the underlying concepts of probability prediction models can help us to clearly recognize what it is that we should be searching for in our statistical analyses of the FCC data. In the next section we describe the approach we have undertaken to discover an efficient partitioning scheme.

2. THE APPROACH

One can hypothesize a host of variables which influence the return probability of a student. Among these are:

- Socio economic variables such as employment status and source and amount of income
- Demographic variables such as age, sex, martial status and family responsiblity
- Academic performance and environment variables such as credit hour accumulated, pass/fail experiences in the previous quarters and other general experience with the college. area of study, etc.
- Personality trait variables such as ambition, commitment to college education/degree and intellectuality.

It is obviously not an easy task to take these myriad of factors into consideration in one comprehensive model. Nor does it seem possible to separate the significant variables from the non significant ones on an a priori basis without making appropriate analyses of actual data. Since exhaustive analyses of this nature are likely to take considerable time even if we have suitable data, a more indirect and expedient approach has been taken first. This report will mainly deal with a discussion of this approach and its results.

In this approach, the past enrollment behavior of a student has itself been sought as a predictive variable of his future reenrollment probability. The rationale for this step has been that, whatever be the variables affecting enrollment the imprint of their sum effect would be left on the past enrollment behavior. Hence the past enrollment pattern suitably quantified, might itself be a predictor. assuming that the variables which operated with the past continue to do so in the future also. Such a predictor may not be as efficient as one using the underlying variables, but this handicap may be more than compensated by virtue of the fact that the necessary data for forming the predictor are readily available.

The enrollment pattern may be quantified by simply observing if the student did or did not enroll in each of the previous four quarters. Enrollment in a particular quarter is indicated by the numeral 1, while non enrollment is indicated by the numeral '0'. With this binary notation, a student who enrolled three quarters ago and not in others in the past four quarters, can have his pattern represented as '0100'. There are sixteen possible permutations over four quarters each giving rise to a particular enrollment pattern. The enrollment pattern represented in this fashion are used as our predictor variable.

In the scheme presented here we ignore students who have not enrolled in any of the previous four quarters.* In other words the permutation '0000' is excluded, leaving us with 15 patterns. The justification for this is the observation that a student who continually absents himself four quarters has less than 10% probability of ever returning to school. Table 1, shows data compiled from a FCC Computer Center Report illustrating this point.** There the number of students dropping out in various quarters is shown. The number of these dropouts who return five or more quarters later (that is after at least four quarters of absence) is shown in the last column of the Table and is seen to be seldom higher than 10%. The percentage which return to school exactly on the fifth quarter after dropout is even less than this, and the same data in Table 1 shows it to be less than 4% for all quarters considered there.

** FCC Computer Center Student Retention Study Report dated September 5, 1972.

* Here and elsewhere in this report the "previous four quarters" refer to the four quarters immediately preceding the one for which enrollment prediction is to be made.

What this means is that in predicting enrollment for a particular quarter affiliate students who have not enrolled in the previous four quarters may be ignored without introducing any large amount of error. As mentioned earlier this has been done in the work reported here. However, it must be noted that the methodology employed can be applied to enrollment pattern over any number of past quarters, so that if the present limits imposed are to be relaxed in future it could be done without undue difficulties.

To represent the above in terms of set theory notations introduced earlier: the set S of total population is considered to be all affiliate students who are yet to pass course X and who have enrolled in at least one of the previous four quarters. This set is partitioned into 15 subsets $S_1, S_2, \ldots S_{15}$ depending on the past enrollment pattern. The following sections of the papers present results of analyses performed to see how well this partitioning scheme could function as a predictive base of future enrollment.

3. DATA PROCESSING AND ANALYSIS

The FCC Computer Center undertook on behalf of the Office of Institutional Research a special processing of FCC grade file to produce data for the analysis. Among other things this task involved forming the sets S and its fifteen subsets (partitions) for each Fall, Winter, Spring and Summer quarters of academic year'72. For each of these subsets the number in the subset as well as the number who enrolled in the quarter under consideration were computed. Table 2 contains this summary data. All the analysis reported below were performed on this summary data.

The first analysis was to see if there were significant differences among the 15 enrollment patterns and the four quarters with regard to reenrollment probabilities. Even a visual examination of Table 2 reveals strong indications of between-pattern differences. Also, summer quarter is evidently different. But any differences between the other three quarters are not apparent to a casual observer. Hence a two-way analysis of variance (with the fifteen enrollment patterns providing one classification and the Fall, Winter and Spring quarters providing the

TABLE 2

RETURN PROBABILITIES BY QUARTER AND PAST ENROLLMENT

Enrollment Pattern Subset	QUARTER FALL 1971 Number in Subset	FALL 1971 Number and % Returned	WINTER 1972 Number in Subset	WINTER 1972 Number and % Returned	SPRING 1972 Number in Subset	SPRING 1972 Number and % Returned	SUMMER 1972 Number in Subset	SUMMER 1972 Number and % Returned
0 0 0 1	500	275 (55.0)	2021	1429 (70.71)	1318	870 (66.01)	965	406 (42.07)
0 0 1 0	506	199 (39.33)	231	35 (15.15)	667	80 (11.99)	501	62 (12.38)
0 0 1 1	534	405 (75.84)	303	232 (76.57)	1545	1181 (76.44)	961	433 (45.06)
0 1 0 0	323	39 (12.07)	431	46 (10.67)	291	11 (15.02)	876	24 (2.74)
0 1 0 1	35	20 (57.14)	315	205 (65.08)	46	27 (58.70)	164	50 (30.49)
0 1 1 0	304	791 (62.83)	141	58 (41.13)	94	19 (20.21)	615	69 (11.22)
0 1 1 1	291	201 (69.07)	486	412 (84.77)	302	230 (76.16)	2315	1036 (44.75)
1 0 0 0	1110	114 (10.27)	686	59 (8.60)	898	37 (4.12)	400	13 (3.25)
1 0 0 1	44	26 (59.09)	214	130 (60.75)	139	89 (64.03)	32	11 (34.37)
1 0 1 0	250	116 (46.40)	44	13 (29.55)	373	87 (23.32)	65	5 (7.69)
1 0 1 1	96	82 (85.42)	94	71 (75.53)	1392	1141 (81.97)	164	66 (40.24)
1 1 0 0	579	173 (29.88)	605	93 (15.37)	214	20 (19.35)	224	10 (4.46)
1 1 0 1	102	73 (71.57)	1453	1188 (81.76)	184	137 (74.46)	84	36 (42.86)
1 1 1 0	1754	1260 (71.84)	258	126 (48.84)	214	65 (30.37)	65	
1 1 1 1	1773	1605 (90.52)	1805	1664 (92.19)	2074	1874 (90.36)	2104	1331 (63.26)

Source: A Computer Center Report dated Feb. 26, 1973 based on grade file.

other) of the data was deemed appropriate to test statistically various hypotheses regarding between-group differences. Since the observational variable is a proportion based on sample sizes that vary from group to group, a two way analysis of variance with unequal cell sizes for proportions was necessary. The proportions in each cell were converted to logit scale by the formula:

$$y_{ij} = \ln \frac{g_{ij} + 1/2}{n_{ij} - g_{ij} + 1/2}$$

where

y_{ij} = the logit scale observation in cell (i,j)*

n_{ij} = number of students in cell (i,j)

g_{ij} = number of students in cell (i,j) who enroll in quarter j

The analysis of variance of the logits is in Table 3.**

---

* cell (i,j) refers to the group of students with enrollment pattern i in quarter j.

** The analysis of variance was performed on the cell OS/360 System using a specially programmed FORTRAN routine called UNOVA 2. For details of the procedure see G. Shedecor "Statistical Methods" 6th Ed, Iowa State Press, pp. 497.

TABLE 3

ANALYSIS OF VARIANCE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARES
Patterns (UNADJ)	.14	6008.97266	429.21216
Quarters (ADJ)	2	120.01675	60.00838
Patterns (ADJ)	14	6054.40625	432.45752
Quarters (UNADJ)	2	74.57837	37.28918
PxQ INTERACTION	28	426.56250*	15.23438
BETWEEN CELLS	44	6555.55078	148.98978

* Interaction highly significant as measured by a χ^2_{28} test. No main effects were therefore tested.

Source: Table 2 data.

The analysis shows that the interaction between quarters and enrollment patterns is itself highly significant as determined by comparing the interaction sum of squares against the percentiles of a χ^2_{28} distribution. This implies that the effect of enrollment patterns is dependent on the quarter so that one cannot talk of a 'quarter effect' or 'enrollment effect' in isolation. More specifically, the existence of interaction shows that the differences between enrollment patterns is itself not constant, but varies from quarter to quarter. Table 2 bears out this point. Compare, for example, the enrollment patterns and "1011" the sole difference between the two being that in the latter the students had also enrolled in the immediately preceding quarter. For Fall '71 the difference between the two patterns in their enrollment probabilities was 0.39 while for Winter and Spring it was .46 and .58 respectively. Consistently similar observations are made when ever two patterns differing only in the last quarter enrollment are compared. The difference between the two patterns is less in Fall than in Spring or Winter.

A simple explanation may be offered to account for the interaction. The quarter immediately preceding Fall is Summer which can be justifiably considered as not the same as the other three quarters of an academic year. It may be hypothesized that enrollment behavior in a Summer quarter is not as strongly related to dropout tendencies in the student as it is in the case of the other 'regular' quarters. Hence the difference between two patterns such as '1010' and '1011' is less marked in Fall than in Spring or Winter.

In comparing enrollment patterns due consideration must therefore be given to difference between Summer and regular quarters. Subject to this qualification, the data in Table 2 strongly suggests that reenrollment probabilities are higher for students who have shown a consistent enrollment behavior in the past.

A rearrangement of the data in Table 2 brings out strikingly clear the various factors affecting the probabilities. Table 4 has been prepared after this rearrangement where we have the data grouped by Fall, Winter and Spring quarters.

TABLE 4

RETURN PROBABILITY ARRANGED BY ENROLLMENT PATTERN OVER REGULAR QUARTERS

Enrollment Pattern Over Previous 3 Regular Quarters	QUARTER							
	FALL '71 Return Prob for those who		WINTER '72 Return Prob for those who		SPRING '72 Return Prob for those who		SUMMER '72 Return Prob for those who	
	did not enroll previous summer	enrolled previous summer	did not enroll previous summer	enrolled previous summer	did not enroll previous summer	enrolled previous summer	did not enroll previous summer	enrolled previous summer
0 0 0	* *	55.00(*)	* *	15.15	* *	5.02	* *	3.25
1 0 0	10.27	59.09	8.60	29.55	4.12	9.35	2.74	4.46
0 1 0	12.07	57.14	10.67	41.13	11.99	20.21	12.38	7.69
1 1 0	29.88	71.57	15.37	48.84	23.32	30.37	11.22	20.45
0 0 1	39.33	75.84	70.71(*)	76.57	66.01(*)	58.70	42.07(*)	34.37
1 0 1	46.40	85.42	60.75	75.53	64.03	74.46	30.49	42.86
0 1 1	62.83	69.07	65.08	84.77	76.44	76.16	45.06	40.24
1 1 1	71.84	90.52	81.76	92.19	81.97	90.36	44.75	63.26

* * Corresponds to the pattern '0000' over four quarters; no data was collected on this group.

* Corresponds to the pattern '0001' over four quarters; this group may include a large number of new enrollees.

Source: Table 2 data.

However, in describing the enrollment pattern we use only the three preceding regular quarters (i.e., ignore the preceding summer quarter). This gives rise to 2^3 or 8 enrollment patterns described by three-digit binary numbers shown in the extreme left hand column of the table. The summer enrollment status is considered in the table by having two columns for each quarter; the first column corresponding to those who did not enroll in the preceding summer and the second column corresponding to those who did. Thus, the entry "4.12" in the first column under Spring quarter against the pattern '100' means that the reenrollment probability in Spring '72 was 0.412 for students who

a) were enrolled in the preceding Spring (in Spring '71) quarter but not in Fall or Winter and

b) were not enrolled in the preceding Summer quarter

The order in which the 8 enrollment patterns are listed is also worth noting. The first row is '000' signifying students who did not enroll in any of the three preceding regular quarters. The next one is '100' when the students had enrolled only three quarters ago (ignoring any Summer quarter enrollment). By hypotheses we expect the group '100' to have higher retention probability than '000'.

Similarily the third pattern in the list, namely '010', is expected to rank higher than '100' if we postulate further that with more recent enrollment experience, the reenrollment probability gets higher. With these two hypotheses as guide, the 8 patterns were arranged to produce an increasing retention probability. The pattern '001' follows '110' in the list with the expectation that though the latter represents more number of quarters enrolled in, the former has more recent enrollment experience.

It is rather very gratifying that the Table 4 data follows the pattern expected thus giving credence to the hypotheses. With only a few exceptions, in all the columns, the probabibilities increase as one goes down the rows. Further, the columns representing Summer enrollment have higher probabilities than the corresponding columns representing non-enrollment in Summer. This is further in line with our hypotheses.

The few exceptions noted mostly occur in the Summer quarter confirming earlier observations that the phenomena affecting Summer enrollment are apparently different.

4. CONCLUSIONS FROM THE ANALYSIS

To summarize the observations made during the analysis, one might conclude the following:

1. Past enrollment behavior of a student does seem to provide a viable base for building predictive models of future enrollment.
2. The more often a student has enrolled in the past, the higher his probability of return is.
3. The more recent a student's enrollment is, the higher his probability of return
4. Enrollment behavior in a Summer quarter is different from those of the other quarters.

The above observations are qualitative. One might be interested in knowing, for example, precisely what quantitatively is the effect of enrollment two quarters ago in a 'regular' quarter on return probability for this quarter. The best way to obtain such quantitative measures is to fit a model incorporating specifically parameters representing these effects. Such a model was constructed and fit to the data. The model construction and results are described below.

A plausible model to represent the effect 'a' of a factor influencing the probability of an event is:

$$p' = p + a(1-p) \qquad -(1)$$

where p' is the probability of the event when the factor is operative and p is the probability of the event where it is not known if the factor is operative or not. 'a' in (1) can be considered as a proportion by which (1-p), the probability of the event not occurring, is reduced. It can also be considered as a conditional probability in the following sense. To make this explanation simpler, assume the event is "reenrollment in Fall '71 quarter", and the factor is "enrollment two quaters ago" (i.e., Spring '71). p is the general reenrollment probability for Fall '71 and 'a' is the probability of reenrollment of students who had enrolled two quarters ago. The reenrollment probability in Fall '71 for students who had enrolled in Spring '71 can then be considered as affected by two forces: one representing the 'attraction' of the Fall '71 quarter to students for reenrollment and the other representing the effects of enrollment 'two quarters ago'. If <u>either</u> of these two forces induce the students to reenroll, we have realized the event. With this as a model of the process, the probability of the event can be easily written down as

$$p' = p+a - ap = p+a\,(1-p)$$

with the additional assumption that the two forces act

statistically independently.*

With the same reasoning it can be generalized that if there are n statistically independent factors acting on a probability p then the combined effect of their forces would be

$$p = 1 - (1-p)\ (1-a_1)\ (1-a_2) \ldots\ (1-a_n) \quad (2)$$

where, a_1, a_2, .. a_n are the individual effects of the factors.

If we define

$$y = \ln\ (1-p')$$

$$m = \ln\ (1-p) \quad (3)$$

and $$b_1 = \ln\ (1-a_1)$$

then (2) can be written conveniently as the additive model

$$y = m + b_1 + b_2 + \ldots b_n \quad (4)$$

m can be considered as the general mean in the model.

It is this additive model using logarithmic transformation (3) which has been employed in constructing the model to fit reenrollment probabilities.

In keeping with the analytical findings (1) - (4) above,

---

* This model is the same as that used for computing reliability of a system with parallel, redundant units.

the parameters considered in the model are:

m = general mean

b_{11} = effect of enrollment 1 quarter ago in a 'regular' quarter

b_{12} = effect of dropout 1 quarter ago in a 'regular' quarter

b_{13} = effect of enrollment 1 quarter ago in the summer quarter

b_{14} = effect of dropout 1 quarter ago in the summer quarter

Similary b_{21}, b_{22}, b_{23}, b_{24}, b_{31}, ..., b_{43}, b_{44} are defined on the enrollment two, three and four quarters ago.

There are 17 parameters including m note that effect of enrollment and dropout in a quarter have been separately introduced as two different effects. The effect of dropout (non-enrollment) in a particular quarter may not just amount to leaving the overall probability p undistrbed but to actually decrease it or otherwise affect it.

With this model one can easily write down the probability p of, say, a student with enrollment pattern '1100' re-enrolling in Winter 72 as follows:

$$\ln (1-p) = m + b_{41} + b_{31} + b_{24} + b_{12}$$

where m = general mean

b_{41} = the effect of enrolling four quarters ago in a 'regular' quarter

b_{31} = the effect of enrolling three quarters ago in a 'regular' quarter

b_{24} = the effect of dropping out two quarters ago in a summer quarter,* and

b_{12} = the effect of dropping out one quarter ago in a regular quarter

The seventeen parameters (m and b's) were estimated by Least Square Method fitting the above model to the 60 observations in Table 2.

Appendix B gives the details of this procedure. As shown there, the model in such that four of the parameters can be arbitrarily set to any value while the remaining thirteen would be then assigned unique values. The parameters b_{44}, b_{34}, b_{24} and b_{14} (representing the effect of dropout in summer quarters) were therefore set to zero and the remaining thirteen derived from the least square criterion

---

* Note that when considering any Winter quarter the summer quarter is two quarters ago.

The values were:

$b_{11} = -0.648887$

$b_{12} = 0.40284$

$b_{13} = -0.840544$

$b_{21} = -0.214826$

$b_{22} = 0.261992$

$b_{23} = -0.485431$

$b_{31} = -0.0705368$

$b_{32} = 0.206000$

$b_{33} = -0.160537$

$b_{41} = -0.54608$

$b_{42} = -0.264895$

$b_{43} = -0.149429$

$m = -0.179431$

Since four of the parameters were given arbitrary values, the above numbers should not be given any absolute meaning, but have significance only relative to each other. How do we measure the effect of "enrollment behavior one quarter ago" from the above data? A valid measure is " $b_{11} - b_{12}$" which compares the effect of enrollment one quarter ago with the effect of dropout in the same quarter. Similarly ($b_{21} - b_{22}$) measures the effect of enrollment behavior two quarters ago etc. These quantities are computed and shown

below:

$$b_{11} - b_{12} = -1.05$$

$$b_{21} - b_{22} = -0.48$$

$$b_{31} - b_{32} = -0.28$$

$$b_{41} - b_{42} = -0.28$$

From model (4) it is clear that a negative represents a force that tends to increase the reenrollment probability . Thus the above data clearly and quantitiatively shows that:

a) enrollment in any of the previous four quarters increases the reenrollment probability

b) the more recent the enrollment, the higher beneficial impact on reenrollment

These, of course, were the conclusions offered earlier, but we have now quantified these hypothesized effects.

The effect of summer quarter enrollment one quarter ago is similarly representd by $(b_{13} - b_{14})$ but since b_{14} was set to zero b_{13} itself is this measure. We therefore have:

$$b_{13} - b_{14} = -0.84$$

$$b_{23} - b_{24} = -0.49$$

$$b_{33} - b_{34} = -0.16$$

$$b_{43} - b_{44} = -0.15$$

Note again the positve effect of enrollment in Summer and how it is stronger for more recent quarters. Further, a comparison with regular quarter enrollment effect shows that the Summer quarter effects are generally smaller. This is in conformance with the fourth conclusion presented earlier in this section.

Finally, we may use the parameter values generated above and see how well they estimate the reenrollment probability. This has been done and presented in Table 5 where for each of the sixty observations (four quarters X fifteen enrollment patterns) the actually observed reenrollment probability as well as those estimated with the model are given. The difference between the two is also indicated. The model is seen to fit the data very well, especially in the regular quarters.

5. IMPLICATIONS FOR CONSTRUCTING THE FORECASTING MODEL

What do all the analyses above imply in regard to our effort to construct a forecasting model for the FCC affiliate population? Essentially, it has shown that is is valid to partition the affiliate population by the past enrollment behavior for prediction pruposes. There are clear differences

TABLE 5

ACTUAL VS ESTIMATED RETURN PROBABILITIES

	PATTERN	ACTUAL	MODEL ESTIMATE	DIFFERENCE
	0001	0.55000	0.55819	-0.00819
	0010	0.39328	0.36438	0.02890
	0011	0.75843	0.72574	0.03268
FALL	0100	0.12074	0.22343	-0.10268
1971	0101	0.57143	0.66493	-0.09350
	0110	0.62829	0.51794	0.11035
	0111	0.69072	0.79200	-0.10128
	1000	0.10270	0.22703	-0.12432
	1001	0.59091	0.66648	-0.07557
	1010	0.46400	0.52018	-0.05618
	1011	0.85417	0.79297	0.06120
	1100	0.29879	0.41377	-0.11498
	1101	0.71569	0.71569	-0.03137
	1110	0.71836	0.63610	0.08226
	1111	0.90525	0.84299	0.06226
	0001	0.70708	0.58820	0.11888
	0010	0.15152	0.27451	-0.12300
	0011	0.76568	0.74656	0.01911
	0100	0.10673	0.10598	0.00075
	0101	0.65079	0.68769	-0.03689
	0110	0.41135	0.44979	-0.03844
WINTER	0111	0.84774	0.80779	0.03994
1972	1000	0.08601	0.11012	-0.02411
	1001	0.60748	0.68913	-0.08166
	1010	0.29545	0.45234	-0.15688
	1011	0.75532	0.80868	-0.05336
	1100	0.15372	0.32511	-0.17139
	1101	0.81762	0.76424	0.05338
	1110	0.48837	0.58465	-0.09628
	1111	0.92188	0.85490	0.06698
	0001	0.66009	0.56448	0.09561
	0010	0.11994	0.22610	-0.10616
	0011	0.76440	0.72965	0.03475
	0100	0.05023	-0.06181	0.11204
	0101	0.58696	0.62907	-0.04212
SPRING	0110	0.20213	0.34088	-0.13875
1972	0111	0.76159	0.76975	-0.00816
	1000	0.04120	0.05887	-0.01767
	1001	0.64029	0.67123	-0.03094
	1010	0.23324	0.41579	-0.18255
	1011	0.81968	0.79592	0.02377
	1100	0.09346	0.19845	-0.10499
	1101	0.74457	0.71999	0.02457
	1110	0.30374	0.50244	-0.19870
	1111	0.90357	0.82618	0.07738

TABLE 5 (Continued)

ACTUAL VS ESTIMATED RETURN PROBABILITIES

	PATTERN	ACTUAL	MODEL ESTIMATE	DIFFERENCE
	0001	0.42073	0.30254	0.11818
	0010	0.12375	-0.23935	0.36310
	0011	0.45057	0.56705	-0.11648
	0100	0.02740	-0.51417	0.54157
	0101	0.30488	0.47105	-0.16617
SUMMER	0110	0.11220	0.06007	0.05212
1972	0111	0.44752	0.67165	-0.22414
	1000	0.03250	-0.71941	0.75191
	1001	0.34375	0.39935	-0.05560
	1010	0.07692	-0.06733	0.14425
	1011	0.40244	0.62715	-0.22471
	1100	0.04464	-0.30401	0.34865
	1101	0.42857	0.54447	-0.11589
	1110	0.20446	0.19054	0.01393
	1111	0.63260	0.71723	-0.08462

between these partitions with regard to the reenrollment probability and these differences can be logically explained.

As mentioned elsewhere, past enrollment behavior may not be the immediate causal factor determining future enrollment of a student, but the analyses leads one to believe that it effectively captures and 'summarizes' the total effect of all the real, underlying factors. An exception may be the affiliate students who do not have sufficiently long past enrollment history to base the prediction on. For example if the student was a new enrollee last quarter, then the sole information in his past history is that he enrolled last quarter. The fact by itself may not be significant enough to tell a lot about his future enrollment.* Possible methods to fortify the model in this respect are discussed in the last section of this paper.

A most attractive feature of this approach which might more than compensate for any of its weaknesses, is the fact that the forecasting model can be operationalized with the FCC

---

* Students newly enrolled last quarter have the pattern "0001". It is interesting to note that in Table 4, it is data pertaning to these points which were out of the general trend observed there. Also, from Table 5, it is seen that the model estimates the probability of "0001" to be about 0.56 in the regular quarters. The proximity to 0.5 can be taken as an indication of the high variability to be expected from this group.

data available today. The master grade file kept by the computer center has information on the complete history of each student since the beginning of the college. From this file, it is a routine data processing problem to select affiliate students who have been in FCC within the past one academic year and then assign them uniquely one of the fifteen enrollment patterns. The generation of Table 2 data from the FCC files of course proves the feasibility of this approach. With the pattern established and the corresponding probability for return in the next quarter, computed, an arithmetic sum of these probabilities would give the expected number of returnees. The actual formula to be used here is of the form (1).

One question arises here. Should we use in (1) the probabilites as they were observed (Table 2) or should we rather use the probabilities as estimates by the model (Table 5). There are some good points about using the model probabilities since they have 'smothered out' random effects in the observed probabilities due to sampling errors and other perturbations. Also with the model we need only thirteen parameters to generate the 60 probabilities whereas, if we were to use the observed probabilities themselves we have effectively a sixty parameter model. Thus, one might prefer the model over the actual observations. However, strictly speaking, this should be done only after the model is further validated

by future data. Thus at this time it is not wise to discard either possibility but must seek to do the necessary validation with more data.

6. FUTHER WORK

In the light shed by the work done so far and reported in this interim technical paper, the following tasks seem to be most relevant to our continued effort to build and implement a forecasting model for freshmen requirement courses. The tasks are not necessarily listed in their chronological order.

1) Validate the model with data similar to Table 2 but pertaining to other quarters, preferably recent ones.

2) Fortify the model, if necessary, with other explanatory variables, enrollment history going back more than four quarters, etc. There is much possibility here. We might compare groups of students comprising the partition '1111' with, say, those comprising the partition '1010' to discover probable explanatory variables. The computer program which generated Table 2, is already capable of providing information such as age, sex, marital status and course history of any individual student. The use of further explanatory variable may be especially beneficial to partitions like '0001' which has a large number of students with little historical enrollment data.

3) Operationalize the model. Systems for data collection analysis and feeding the model have to be set up. The system should be capable of forecasting more than one quarter hence and also capable of accepting data on 'control variables' (such as number of new enrollees to be taken into the school in the coming quarters) and integrate them meaningfully into the forecast. The general mathematics for accomplishing this has been established (see Appendix A) but remains to be 'particularized' to the final forecasting model we will be coming up with.

APPENDIX A

OPERATIONALIZING PREDICTIVE MODELS

While discovering predictive variables is often the major problem in a forecasting task, designing an operational system to use the prediction scheme is of no less importance. From the operational point of view there are three aspects to be considered. First, one must determine the means for measuring the predictor variables from available data and design the information system which would regularly supply the necessary data. This task largely depends on the second and central aspect of the problem, namely constructing the mathematics which would let one take the data and transform it into the predictor variable and then into the forecasts for one or more future periods. The third aspect to be considered is that of monitoring the system to detect those changes which would oblige us to modify the forecasting scheme or shifts in model parameters.

The forecasting models of interest in this study proceed by partitioning the set S of affiliate student population into subsets S_1, S_2, ... S_m and computing the probability r_1, r_2, ... r_m to be associated with these sub sets. The forecast is then given by $r_1 s_1 + r_2 s_2 + \ldots + r_m s_m$ where s_i = number of student in S_i. Given S_1, S_2, ...S_m the forecasting task is simple. But the availability of this inform-

ation cannot be always taken for granted. The forecast may need to be made at a time when the data required to compute one or more of the S_i's are not yet available. This is invariably the case when one is trying to forecast for a number of future periods using a model which needs the S_i variables for the immediately preceding period. The model using past enrollment pattern as a predictive variable discused in the text is a good example. To predict the number of returnees from the affiliate student population for quarter i, we need to classify each student in one of fifteen categories based on his enrollment history in the four quarters (i-4), (i-3), (i-2) and (i-1). Therefore, if we are required to forecast for quarter (i+1) as well as quarter i at the same time, strictly speaking we will need enrollment data for quarter i which of course would not be available to us.

A natural way to deal with this situation is to find a means for forecasting the predictive variable $s_1, \ldots s_m$ for period (i+1) from the knowledge of these variables for the period i. In other words, if $s_1(i), s_2(i), \ldots s_m(i)$ stand for the predictor variables in period (i), then what we need is a set of transform $f_1, f_2, \ldots f_m$ such that:

$$s_1(i+1) = f_1(s_1(i), s_2(i), \ldots s_m(i))$$

$$s_2(i+1) = f_2(s_1(i), s_2(i), \ldots s_m(i))$$

$$s_m(i+1) = f_m(s_1(i), s_2(i), \ldots s_m(i))$$

The above equations, may be called the 'System Dynamics Equations' since in a sense they describe the movement of the system from one period to the next. Quite often in order to completely specify the dynamics, it is necessary to introduce variables other than $s_1, \ldots s_m$. For example, to specify, the number of students in the category '0001' for quarter (i+1) we must know the new enrollees to be admitted into the school in the quarter i. Variables like these needed to completely describe the system dynamics may be called "input variables" and denoted by $(e_1, e_2, \ldots e_n)$. Some of the input variables may be determined by decision makers, while, others may be purely dependent on the 'state of nature'. Including the input variables, the system dynamics equations assume the vector form

$$\underline{s}(i+1) = \underline{f}(\underline{s}(i), \underline{e})$$

Where $\underline{s}(i+1)$ and $\underline{s}(i)$ are Mx1 vectors, $\underline{f}$ is a Mx1 vector function and $\underline{e}$ is a nx1 vector. One of the prime needs for forecasting for a number of periods in therefore the ability to predict the input variables $\underline{e}$ also.

It is intersting to note here that the model using four quarter enrollment history as a predictor variable has an associated set of system dynamics equations that are quite simple. The only imput variable needed for completing the dynamics equations is the number of fresh entrants into the

affiliate population in the quarters of interest. This variable is to a certain extent a control variable and should not prove difficult to assign values to.

Designing an information system to support the predictive model with necessary basic data is not very difficult either. What would essentially be needed is a computer program to go against the current grade file to classify all affiliate students (or at least those who have enrolled in one or more quarters in the past one year) into one of the fifteen groups and count the number of students in each group. Such a computer program exists even now and it should be possible to make it a 'Production program' for regular use with little or no modification. With this data and the estimates of fresh inputs expected over the future quarters, forecasting over any number of quarters can be made.

The third aspect mentioned also deserves attention. The forecasting method is only as good as the parameters used. In context of our particular model, the parameters are the return probabilities for the various classifications. It is necessary that a continuous check be made with current data to see if the parameter values used are indeed valid. For key parameters, quality control charts or other statistical/graphical techniques may be used to do this monitoring.

Again it is feasible to integrate this aspect with the total forecasting system so that as forecasts are made and compared with the actual, the current parameters values are evaluated and tested for possible significant shifts.

One of the tasks in the next phase of the study would be to build such an integrated forecasting system out of the basic model developed in this paper.

APPENDIX B

FITTING AN ADDITIVE MODEL TO LOGARITHM OF REENROLLMENT PROBABILITIES

If p is the proportion of students with a particular past enrollmet pattern who reenroll in quarter i, then the model hypothesizes that:

$$\ln(1-p) = m + b_{4i} + b_{3j} + b_{2k} + b_{1l} + e$$

where each of the indices i,j,k,l take one of the four values 1,2,3,4 according to the following rule stated for b_{4i}.

(1) i takes the value 1 (i.e., b_{4i} is the variable b_{41}) if the students had enrolled 4 quarters ago in a regular (non-summer) quarter.

(2) i takes the value 2 (i.e., b_{4i} represents b_{42}) if the students had not enrolled four quarters ago in a regular quarter.

(3) i takes the value 3 (i.e., b_{4i} is b_{43}) if the students had enrolled four quarters ago which was a summer quarter.

(4) i is 4 (i.e., b_{4i} is b_{44}) if the students had not enrolled four quarters ago in a summer quarter.

The above rule is also applicable to b_{3j} with the difference that the enrollment behavior three quarters ago (instead of four quarters ago) is considered to determine the appropriate index. Similarly b_{2k} and b_1 are determined based on enrollment two and one quarter prior to quarter i.

As an illustration, if we are condidering the reenrollment probabilty p of students with enrollment patter '1011' in a Summer quarter, then the model suggests that ln (1-p) is given by

$$\ln(1-p) = m + b_{43} + b_{32} + b_{21} + b_{11} \qquad -(1)$$

the random variation due to e having been ignored.

We can similarly write down the expression for any of the 15 enrollment patterns and four quarters for which data is available.

In this model there are seventeen parameters including m and the sixteen b variables. In the least square method, these parameters are estimated from the data so as to minimize the sum of squares of deviation of the observed ln (1-p) and estimated ln (1-p) for the sixty observations we have. Since p is given by (1) in terms of the parameters, this method amounts to minimizing:

$$L = \sum_{1}^{60} (p - m - b_{4i} - b_3 - b_{2k} - b_{1L})^2$$

with respect to the m, b parameters. The summation is over the sixty observations, the appropriate parameters b_{4i}, b_{3j}, b_{2k}, b_{1L} being determined according to the rules given before for each observation. The parameter values minimizing L are obtained by solving the equations resulting from setting the the partial derivative of L with respect to each parameter to zero. The equations are (called normal equations) all linear and are exhibited in Table B-1 in matrix form.

The 17 by 17 matrix is singular and in fact has rank 13 as can be seen by the fact that $(b_{21} + b_{22} + b_{23} + b_{24}) = m = (b_{11} + .. b_{14})$ This means, in terms of parameters estimation that we can give four parameters any arbitrary values and then determine the rest. Acoordingly b_{14}, b_{24}, b_{34}, b_{44} (representing effect of dropout in Summer quarters) were all set to zero and removed from the equations. Equations 4,8, 12, and 16 coresponding to these variables were also dropped. The remaining 13x13 matrix is non-singular and was solved using a CALL OS/360 routine called SIMQ available in FCC library. The estimated parameter values are:

$b_{11} = -.648887$	$b_{31} = -.070537$
$b_{12} = .40284$	$b_{32} = .206003$
$b_{13} = .840544$	$b_{33} = -.160537$
$b_{21} = -.214826$	$b_{41} = -.54608$
$b_{22} = .261998$	$b_{42} = -.264895$
$b_{23} = -.485431$	$b_{43} = -.149429$

$M = -.179431$